Practical Assessment, Research, and Evaluation

Volume 16 *Volume 16, 2011* Article 17

2011

Reliability of Grading High School Work in English

Hunter M. Brimi

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 16, Number 17, November 2011 ISSN 1531-7714

Reliability of Grading High School Work in English

Hunter M. Brimi
Farragut High School, Knoxville, TN

This research replicates the work of Starch and Elliot (1912) by examining the reliability of the grading by English teachers in a single school district. Ninety high school teachers graded the same student paper following professional development sessions in which they were trained to use NWREL's "6+1 Traits of Writing." These participants had been instructed to construct a 100-point rubric, assigning point values to each trait (though not all complied with this request). To evaluate the reliability in grading, data were analyzed for teachers reporting scores on a 100-point scale. Of the 73 participants who graded on a 100-point scale, the scores ranged from 50 to 96. Analysis suggests that many of these teachers are proficient at assessing student writing, many are unaware of or simply resistant to research suggestions for writing assessment, and many show signs of being "assessment illiterate" (Stiggins, 1995).

Nearly a hundred years ago, Daniel Starch and Edward Elliot (1912) confirmed what legions of students already suspected: Teachers *give* grades as much as students earn them. By distributing the same two English papers to 200 teachers, they found that different readers assigned different grades to the same work. These researchers wrote:

> *The reliability of the school's estimate of the accomplishment and progress of pupils is of large practical importance. For, after all, the marks or grades attached to a pupil's work are the tangible measure of the result of his attainments, and constitute the chief basis for the determination of essential administrative problems of the school, such as transfer promotion, retardation, elimination and admission to higher institutions; to say nothing of the problem of the influence of these marks or grades upon the moral attitude of the pupil toward the school, education, and even life. (p. 442)*

In the century since Starch and Elliot's publication of "Reliability of the Grading of High-School Work in English," little has changed in our (and the students') views of grades. They still present quantifiable evidence of student achievement, they still help open the doors to higher education, and they still, too frequently, determine how students view themselves.

What has changed is the magnitude of consequences connected to grades. For example, many states reward students for their good grades (and qualifying college entrance exam scores) with scholarships to state colleges and universities. As competition for these scholarships has increased, more questions arise as to the meaning of a grade. If students must maintain a "B" average to earn a scholarship, then what does this mean about the quality of their work? Is a "B" the same on one end of the state as it is on the other? Is it even the same within one school district, or on a single school hallway?

In response to these questions, more states have mandated a fixed grading scale to mollify those who believe that a percentage represents truth. Pity the poor student with an 84% in English in a school system where a B starts at 85%: His counterpart in another part of the state may attain the grail-like B with the same percentage; his counterpart may have an advantage in earning a state scholarship. As of this writing, 19 states (including the state in which this research occurred) have legislated uniform grading scales. Politicians, parents, and other laypeople may view this as insurance of equity in determining who earns qualifying grade-point averages for the purposes of college admissions and scholarship recognition.

But do grading scales affect the teachers' perception of the work? And even if they do not, does this certify uniform grading standards? Even in 1912, the answers to both questions would have clearly been "No."

Starch and Elliot recognized that their participants used disparate grading scales: of the 142 responses they evaluated, 51 teachers worked in schools where 70 was the benchmark for a passing score; 91 taught where 75 was the standard. In their work, these researchers found little difference in the median scores given by those teachers whose passing mark was 70 and the scores given by teachers working with a passing score of 75 (p. 450-451).

Yet, Starch and Elliot found a "startling" range of scores overall: "It is almost shocking to a mind of more than ordinary exactness to find that the range of marks given by different teachers to the same paper may be as large as 35 or 40 points" (p. 454). To illustrate the effect of this range, the researchers commented on the disparity between the score of paper B as given by the student's actual teacher and that given by the other teachers. Whereas this paper achieved a passing score when graded by the student's teacher, 22 graders gave the same paper a failing mark. On the other hand, the students' actual teachers granted grades (80% and 75%) that were *lower* than the median of the respondents (87.2% and 78.8%), indicating that they may have been "tougher" graders (p.454). As Starch and Elliot note, "Therefore, it may be easily reasoned that the promotion or retardation of a pupil depends to a considerable extent upon the subjective estimate of his teacher" (p. 454). Additionally, the researchers found that teachers from "small" schools (i.e., school populations of 150 or less) graded "more liberally" (p. 457). Such discrepancies were perhaps tolerable at the time.

A century later, though, we live in an era of increased standardization. Does this mean that the subjective elements of grading have subsided? This general query guides this research: With training on using a set of performance indicators (NWREL's "6+1 Traits of Writing"), would teachers differ as greatly as their 1912 predecessors in their scoring of exact copies of the same paper?

The purpose of this study is to assess the grading reliability of English teachers within one school district. Would teachers working in the same school district, having received the same training on specific performance indicators, assign statistically similar grades to the same paper?

REVIEW OF THE LITERATURE

To understand the task the participants undertook, we must consider the condition of assessment in American education. Stiggins (1991, 1995) has asserted that, when it comes to assessment, educators are largely "illiterate." That is, they do not fully comprehend their purposes in creating assessments, nor do they understand how to best formulate assignments for assessment. Too often these teachers seek simply to emulate formats of current educational trends in the area.

In the 1980s these trends favored multiple-choice assessments. Consequently, teachers tended to copy this format in their own day-to-day assessments without ample knowledge on how to do so in a way that genuinely gauged student learning (Stiggins, 1999). By the early 1990s, though, performance-based and alternate assessments became popular. Educators floundered in attempts to adapt in their own assessments due to a lack of training, a situation stemming from failures in colleges of education and school districts (Stiggins, 1995). Too many teachers failed to create assessments that presented students the chance to demonstrate understanding while also illuminating facets in which students were deficient.

According to Stiggins (1999), there have been several impediments to progress in overcoming the assessment illiteracy that has hindered educators. While teachers have, at times, realized their own limitations in the field of assessment, they have been largely incapable or simply unwilling to make changes to their current practices (Stiggins & Bridgeford, 1985). Indeed, Stiggins (1986) later commented on the stark discrepancy between recommended practices and what actually occurs in classrooms. In terms of writing assessment, Hillocks (2006) attributed this reality to a combination of teachers' ignorance of research and/or their indifference towards research.

Aside from stubbornness, teachers face other barriers to improving their ability to assess student work. For one, they often lack the time or administrative support essential for this type of professional development. Teachers also have demonstrated that they use assessment as a motivator, not as an instrument for student learning. That is, they find that students are more willing to pay attention, to "learn," if they know that class material will be tested (Kahn, 2000; Stiggins, 1999). Researchers also believe that standardized tests have exacerbated the teachers' problems with daily assessment (Hillocks, 2006; Stiggins, 1999).

A more difficult obstacle to improving assessment, however, lies in the teachers' own content knowledge, or lack thereof. Stiggins (1999) questioned the ability of teachers to assess content that they had not themselves "mastered." Hillocks (2006) also recognized this barrier as it pertains to writing assessment and instruction. Although they have learned several methods for assessment in the past three decades (holistic scoring, primary trait analysis, analytic scales, etc.), teachers misuse these methods if they lack requisite knowledge of writing. Hillocks (2006) claimed that teachers suffered from a lack of in-depth training from their colleges of education and school districts. Consequently, he argued that much of the teachers' knowledge of writing has derived from the requirements of state writing assessments. These assessments, though, do not encourage sound argumentation, nor do they reward writing that falls outside of a prescribed formula (Hillocks, 2005).

This prescribed formula most frequently is the five-paragraph theme, a convention that NCTE researcher Janet Emig (1971) called the "Fifty-Star Theme" due to its ubiquitous presence in American high schools (p. 98). Furthermore, Emig noted that as teachers belabored the merits of this writing formula, students held "inward cynicism and hostility" toward their writing instruction because of its lack of relevance to real-world applications (p. 93). She ultimately condemned the writing instruction of her day thusly:

> Much of the teaching of compositions in American high schools is essentially a neurotic activity. There is little evidence, for example, that the persistent pointing out of specific errors in student themes leads to the elimination of these errors, yet teachers expend much of their energy in this futile and unrewarding exercise. (p. 99)

Despite Emig's animadversions toward this type of instruction, the use of the five-paragraph theme pervades writing instruction even today, as Hillocks (2002, 2005) argued.

Hillocks (2005) further condemned standardized writing assessment because it encourages writers to concentrate on "form" not "content." In citing his disapproval of timed writing assessment, Hillocks even demonstrated that the standards can mislead teachers into teaching flawed constructs (p. 246). He warned that teachers of composition must be cognizant of avoiding writing instruction that aims simply to duplicate forms and techniques in order to facilitate student achievement on standardized assessments.

For example, Scherff and Piazza (2005) indicated that the use of product-oriented testing in Florida has led to a decrease in the use of the writing process. In their survey of high school students, the researchers found that teachers were asking students to write more than in the past, but that "little class time was used for writing conferences or peer review resulting in revision of multiple drafts" (p. 293). This illustrates how the test may not only influence *what* is taught, but *how* it is taught as well. Moreover, these findings could confirm that the teachers lacked much knowledge of writing instruction even before teaching for the state writing assessment.

Kahn (2000) similarly found a dearth of knowledge amongst her research subjects, a group of 15 high school English teachers that collaborated on writing instruction for sophomore English classes in their school. These teachers had devised a rubric for grading the five paragraph essays of their students, and, to prove the reliability and lack of bias amongst graders, the teachers regularly exchanged papers so that they would not grade those written by their own students. As admirable as this goal may seem, Kahn decried the utter inadequacy of the teachers' rubric commenting, "This writing assessment appears to focus more on whether students can write a composition in a fairly rigid five-paragraph form than on the overall persuasiveness or quality of the ideas and support presented" (p. 280-281).

Kahn's research also exposed another possible complication in the process of curtailing assessment illiteracy: the teachers' reasons for clinging to their traditional pedagogy and assessment modes. The teachers felt that students were more attentive and well-behaved when presented with subject matter over which they would be tested. And in terms of tests, these teachers focused on content knowledge, not the invention and expression of ideas that mark high-quality writing. In other words, they believed that students viewed discussions and the exchange of ideas as unimportant since they could not be memorized and used for testing purposes. Incredibly, these teachers developed assessment (and instruction) not to directly enhance student learning, but to more effectively maintain orderly classrooms.

More researchers and commentators, though, have espoused assessment programs that focus more on the needs of students, not teachers. In writing about Curriculum Based Assessment (CBA), Charles Hargis (1995) advocated using tests as an evaluation of instruction, not students. Specifically, he urged teachers to use assessment scores to gauge whether the

curriculum matched the students' instructional level, as articulated by Betts (1946). Teachers using CBA, then, would seek primarily to match their instruction to the students' needs in order to ensure their success.

Stiggins (2007) espoused assessment theories that not only worked toward student success, but that were also marked by a sense of humanity. He wrote that assessment should reflect learning goals that teachers have shared with their students. The teachers also should share examples of quality work, according to Stiggins. Ultimately, the students should be able to track their own progress, their own triumphs, and their own shortcomings. Stiggins also redefined the notion of test validity, stating that tests are invalid if they result in students quitting due to their difficulty. He added that assessments can only be valid if test-makers seriously account for the students' "emotional" responses to the work asked of them.

The NCTE's Conference on College Composition and Communication (CCCC) has promoted similarly student-friendly goals for writing assessment (2006). In its most recent policy statement on writing assessment, the CCCC has made the following germane dictums:

> The methods and criteria that readers use to assess writing should be locally developed, deriving from the particular context and purposes for the writing being addressed…Best assessment practice clearly communicates what is valued and expected, and does not distort the nature of writing or writing practices…Best assessment practice enables students to demonstrate what they do well in writing. (2006)

These principles appear to contradict several assumptions currently made about writing assessment. First is the notion that all (or at least most) writing situations require the same evaluative treatment. Second, NCTE recommends that teachers should avoid primarily evaluating writing mechanics and structure at the expense of other elements of writing such as invention. Third, teachers should resist a punitive attitude toward writing assessment. That is, they should avoid merely cataloguing the mistakes of their students in order to justify grade deductions; they should reward students for their successes.

Finally, with these precepts in mind, consider the views of Peter Elbow (1998) on the problematic nature of writing for teachers. Elbow wrote:

> When you write for a teacher you are usually swimming against the stream of natural communication. The natural direction of communication is to explain what you understand to someone who doesn't understand it. But in writing an essay for a teacher your task is usually to explain what you are still engaged in trying to understand to someone who understands it better. (p. 219)

In this sense, the teachers' presumably abundant knowledge on assigned writing *topics* creates an unnatural writing situation. Furthermore, this could lead to discordant assessment if the teachers allow their own experience and familiarity with a topic to affect their perceptions of the students' writing.

Perhaps Elbow (1998) made a more pertinent observation in terms of assessment when he illustrated another problem in the teacher-student exchange:

> *I can't really read for enjoyment when I'm not free to stop reading. I can't just sit back and be enlightened or entertained. I must look for weaknesses and mistakes. Inevitably I improve. But students don't improve with me. That is, each year I get better at finding weaknesses and mistakes, but each new batch of students is just as unskilled as last year's batch. Thus, every year I find more mistakes and weaknesses per page. (How could I not believe that students get worse every year?)* (p. 224)

If Elbow's experience reflects the experience of other teachers of writing, then the reliability of a teacher's grades would be compromised from year to year. Worse still, the teacher may develop a propensity to harshly delineate her students' most picayune of mistakes, a trait that could detrimentally affect the students' will to write. In this sense, teachers should act less as evaluators, and more as guides for their students (Elbow, 2000; Hairston, 1982; Hillocks, 1986;).

METHOD

Procedures

Unlike the research of Starch and Elliot (1912) that used data collected from 200 schools in the North Central Association, this research considers the reliability of grading by English teachers within a single school district. Furthermore, these teachers had been trained to use a specific grading system, NWREL's 6 +1 Traits of Writing, whereas the teachers in the Starch and Elliot study did not employ any specific performance indicators (See Appendix A).

In the 2007-2008 academic year, 90 ninth and tenth-grade teachers representing 12 schools were trained to use 6+1 as a teaching and assessment tool. This training was mandatory for all teachers as the school district's Language Arts Department had chosen to adopt the 6+1 model to assess writing at all grade levels. This model is comprised of "Ideas," "Organization," "Voice," "Word Choice," "Sentence Fluency," "Conventions," and "Presentation" (Culham, 1995). The school district's writing coach (a former teacher who had selected the 6 +1 instrument after attending a week-long workshop on the model) worked with these teachers for two days (16 hours total) during a single summer. The ninth-grade teachers had received an additional three hours of training the previous year and in the fall of 2007; the tenth-grade teachers, two hours in the fall. The first day of summer sessions focused on defining the traits; day two involved work on grading essays using these traits. Sessions during the school year were used to reinforce the lessons of the summer and entailed discussion of problems in using the system, direction for using the traits in classroom instruction, and guidance for finding additional resources for instruction.

In the spring of 2008, these same teachers attended follow-up sessions at their respective schools and were presented with copies of the same student paper to grade (the paper, "Why Abortion Should Be Illegal," can be found in Appendix B). I procured this paper from a former student who willingly and anonymously volunteered her essay. I chose this particular essay because it included several strong points as well as several flaws. (I had scored the paper at 83%--high "C" according to the district's grading scale). Furthermore, I believed most teachers in the district would feel most comfortable grading argumentative work due to the state's focus on persuasive writing in the eleventh grade writing assessment. I acknowledge, too, that the graders' individual views on this controversial issue could have skewed their judgment and contributed to the wide range of scores. I also argue that as a matter of professional ethics, personal viewpoints should not influence student grades.

The participants were instructed to devise a rubric using the 6+1 traits and to score the essay on a 100-point scale using this rubric. In effect, they were asked to use primary traits assessment, a method that requires graders to analyze different aspects of writing individually. The graders were allowed as much time as needed to complete this task and were asked not to consult with other teachers during the process. They were also informed that the paper was a final draft of a research paper.

The graded essays were collected and coded according to the schools where their respective graders taught. To assure anonymity, no data was recorded regarding the teachers' gender, experience, or educational background.

Population

The participants taught in a school district that served over 400,000 residents, including 17,000 who were registered in public school. Of the population at-large, 86.5% of residents 25 and older had attained a high school diploma or equivalency; 32.3 had at least bachelor's degree. Fourteen and a half percent of families with children under 18 lived below the poverty level. The mean household income in this district was $62,153 while the median was $44,961 (U.S. Census Bureau, 2007).

Of the schools represented in this research, at the time only four were in "Good Standing," according to the guidelines of NCLB. The district had a graduation rate of 79.2%. Of these graduates, 61% scored 21 or higher on the ACT, with an average score of 22.0 (the state average was 20.6). In Reading and Language Arts plus Writing, 90 % of this district's high school students scored "Proficient" or "Advanced" compared to 91% for the state as a whole (Tennessee Department of Education, 2007).

In 2007, the district had a K-12 expenditure of $7,732 per student. The K-12 population consisted of 80.3% Caucasian students, 14.7% African-American, 2.8% Hispanic, 1.9% Asian/Pacific Islander, and .3% Native American/Alaskan (Tennessee Department of Education, 2007).

Given the purpose of this study---to determine reliability in grading across the school district and within individual schools---it is also necessary to know the grading scale upon which these participants based their evaluations. This school district used the following state-mandated grading scale: A percentage score from 93 to 100 equals an "A." A score from 85 to 92 is considered to be in the "B" range. The range for "C" scores begins at 75 and stops at 84. A score from 70 to 74 equals a "D." Any score below 70 represents a failing grade.

RESULTS

The purpose of this study was to gauge the consistency of grades given to a single paper by a variety

of teachers from one school district. Teachers had been trained to use NWREL's 6+1 Traits of Writing as an instructional and assessment tool prior to scoring the student paper. Of the 90 participants, 89 returned graded papers (one teacher simply refused to participate). Three papers were discarded for the following reasons: The first provided only two letter grades (a "C" based on the author's use of "you" and a "B" based on the writer's content). The second recorded illegible trait scores along with a final score of "B." The third only provided a letter grade of "C." Of the remaining 86 participants, 73 adhered to the request to give a numerical grade based on a 100-point scale.

Distribution of Scores

Within the population of 73, the range of scores was 46 points (high of 96; low of 50). These teachers gave the paper a mean score of 81.1599 (See Table 1). These participants also assigned a total of 30 different scores. The most common scores fell at minimum/maximum scores for a letter grade or at numbers divisible by ten, including five scores of 70; eight of 75; six scores of 80; six scores of 84; six scores of 90; and five scores of 93 (see page 5 for the school district's grading scale). As for letter grades, ten participants scored the paper as an "A," 18 assigned a "B," 30 marked the paper a "C," nine gave a "D," and six graded it as an "F."

Table 1: Descriptive Statistics for Grades on 100-Point Scale

	N	Minimum	Maximum	Mean	Std. Deviation
Final Grade	73	50.00	96.00	81.1599	9.55938

This table shows the grade range, the average grade, and the standard deviation of grades assigned by participants. See Appendix C for a complete list of the frequency of final grades on a 100-point scale.

CONCLUSIONS AND DISCUSSION

This study's research question considered the issue of grade reliability amongst English teachers in a single school district. Specifically, I asked: Would teachers across the district, having received the same training, assign the same paper grades that lie within a range similar to the ranges shown in the Starch and Elliot (1912) study?

The data show the answer to be fairly plain: Despite several sessions of training in using the same grading methods, these participants awarded final scores that were as discrepant as those recorded in the Starch and Elliot (1912) study. In that study, the ranges of the grades on the two papers were 37 and 44. The researchers noted that they believed the wide ranges were due "to a small extent, to the differences in method of teaching and in the emphasis and importance placed by different teachers on different aspects of English" (p. 454). In this study, the range of scores for the single paper within the school district was 46.

This data leads to the following conclusions:

1. English teachers within this district evaluate writing differently.
2. As a result, a wide range of scores exist for the same quality of work.

Any discussion of these conclusions has to begin with the question, "Why?" Why did the same teachers give vastly different scores to the same paper? Why did these teachers have such different impressions of this writer's proficiency in the 6+1 Traits of Writing?

For one, these participants may be very much like the teachers Hillocks (2006) and Stiggins & Bridgeford (1985) discussed. That is, they are either ignorant of current research and practices in grading writing, or perhaps they just do not care to change their views on writing assessment. Evidence of the latter came from outside the quantifiable data. For instance, many teachers failed to comply with the requests made in terms of developing a 100 point rubric. Also, some teachers made few to no comments or even marks on the papers, but instead just produced a grade. Other teachers relied on methods that predated the district's mandate that teachers use the 6+1 traits, such as putting Harbrace numbers above conventions (grammar) errors.

The data also may confirm Hillocks' (2006) belief that many teachers lack preparation to teach composition at a level beyond the basic requirements of state assessments. The range of scores suggests that the teachers may not understand what solid "Ideas," or strong "Word Choice," or effective "Sentence Fluency" entail. Furthermore, some of the teachers appeared to cling to the earmarks of the "five-paragraph" theme. Several teachers penalized the student for using a delayed thesis, a technique that is discouraged by the state's writing assessment, but that is strategically sound when

writing about a controversial topic (in this case, abortion). Other teachers criticized the paper for using the word "you" in the opening hypothetical situation. On the other hand, despite the paper's obvious flaws in using MLA documentation, most teachers did not penalize the student's presentation score.

Also, some teachers seemed to focus on what they could mark wrong. Indeed, those who graded the paper most unfavorably made few to any remarks on any strengths of the paper. They made copious comments and clearly marked deductions for grammar errors. Some of the scores given for "Conventions" (18%, for example) would seem to indicate that the paper was nearly incomprehensible due to grammar flaws.

The teachers also appeared to lack a clear understanding of how to derive a final grade after assessing each trait. As previously discussed, some teachers gave very little information as to how they arrived at their final scores. One teacher simply put letter grades, not numbers, next to each trait. Even those who created rubrics frequently did a disservice to the student in the way they assigned credit. For instance, one participant placed the numbers "5" (out of 6) and "12" (out of 14) in her assessment of "Ideas." The "5" should indicate that the writer's ideas were strong, but a 12/14 is approximately an 85%, the lowest score in the "B" range. The same teacher assigned 9/14 (64%) to the score of "4" ("Effective") for more than one trait. If the student challenged this grading, could the teacher adequately justify the score by saying, "Yes, your word choice was effective, and that's six points below passing"? A statement such as this might mark a teacher as "assessment illiterate" (Stiggins, 1995).

On the other hand, perhaps these punitive rubrics were not the result of ignorance, but of careful calculations to keep grades low. Kohn (1999) wrote of teachers who felt that if their students were not failing, then they, as teachers, were not doing their jobs. Some teachers, too, may feel pressure to be "hard" or, at least, "challenging." They may feel that their colleagues will view them as "weak" or unwilling to "uphold standards." In this study, teachers from different schools may have felt the need to "prove themselves" to the county supervisor and the writing coach as proponents of academic rigor.

This raises the question: should a paper receive the same grade regardless of where it was written and for whom it was written? After all, audience remains a key component of rhetoric. To be effective communicators, writers must be aware of their audiences' disposition and knowledge. The unwanted effect of this could be teacher prejudice for or against a student's argument. Maybe those who were "Pro-Choice" could not separate their disagreement with the writer's position from their judgment of the writing. Or "anti-abortion" participants may have been more accepting of the argument and subconsciously (or consciously) more generous in their assessment.

A more digestible effect of these different audiences from different communities may lie in the teachers' understandings of their students' strengths, weaknesses, and ultimately, their appropriate level of instruction. That is, teachers accustomed to grading poorly written papers, may have viewed the errors in this paper as negligible. The paper, therefore, may have exceeded the standards of those teachers' individual schools. On the other hand, other teachers may have viewed the paper with higher standards for the elements of a "C" or "B" level paper. Of such standardization, Hargis (1995) wrote:

> All too often individual differences in learning ability are viewed as curable maladies. However, our attempts to cure them produce more casualties. We make the misguided attempt to force children to perform up to grade level standards…By the time primary-age children reach high school, the range [of academic capability] exceeds five years. (p. 6)

In essence, Hargis argued that teachers should strive to help students make attainable progress, acknowledging that they may not reach the same standards that their classmates will reach or even exceed. The grade, therefore, should not be based upon comparisons to the work of more advanced or less proficient students.

Teachers in this study did not know the academic level of the student whose paper they were grading, and consequently, their role as an audience may have been suited to the writing of students of lesser ability. If this were the case, then the small grammar errors that some teachers consistently marked may not have been such gross examples of poor language. The ability of this writer to combine sentences in a variety of structures may have outweighed shortcomings such as misplaced commas, incorrect citations, or a sentence fragment…or the seemingly unforgivable use of second-person that one teacher saw as primary grounds to give the paper a failing grade.

This lack of knowledge of the student may mark a weakness of the study. After all, the participants did not know about the instruction the student had received or the readings she had studied before this assignment. The

students' aforementioned use of second-person had been influenced by an essay on academic pressure and cheating that my class had read. In this essay, the author had effectively used a second-person hypothetical situation to connect her audience to a specific circumstance, just as the student tried to place readers in the position of the young girl in the abortion clinic.

Regardless of their knowledge of the student, though, the teachers showed several shortcomings in their understanding of writing. One participant wanted to know whether the paper was a research paper or an argumentative essay, apparently unaware that an argument can benefit from research. Again, others were confounded by the delayed thesis. Does this mean that these teachers only teach their students to place a thesis at the end of the first paragraph, that the thesis should have three points, that these three points should dictate the topics of the three body paragraphs, that the five-paragraph theme is the only form of writing that high school students should know? If so, then these teachers are propagating a puerile approach to writing that endangers their students' growth.

So what should students expect from their teachers in terms of assessment? Should an "A" in Mrs. Smith's class be an "A" in Mr. Jones's? Should a student who is below grade level be held to the same standard as one who is above grade level? While this study does not aim to answer these questions, it does indicate that an "A" in one class may not be an "A" in another class or at another school. And the Advanced Placement student who received an 83% on her abortion argument in my class might expect anything from an "F" to an "A," depending on who grades the paper and what the grader knows about writing and assessment.

There are several large-scale implications of this subjectivity in grading. For one, grades help determine which students colleges admit and which students receive scholarships. Universities may wish to rely on their own assessments of a student's writing sample to evaluate the student and admissions officers may be wise to view English grades somewhat skeptically. Scholarships, especially those funded by state lotteries and based largely on grade-point average, are a more troubling matter. If students qualify for such scholarships based on inflated grades, then their college experiences may be marked by futility and the funding effectively rendered a lost investment when students fail to earn a degree.

This type of disparity in grading may also lead to teacher-shopping within a school. As teachers garner reputations as easy or hard graders, students (and parents) may increasingly pressure administrators and guidance counselors for preferable placement. This results in a phenomenon in which one teacher has class sizes significantly lower than colleagues holding reputations as "friendly" graders (as I have witnessed in my own professional experience).

Grading subjectivity may also result in another response: increased training for and emphasis on standardization—not just of assessment, but of writing assignments as well. In the district in which this study took place, common rubrics have already been devised for assessing research papers and student presentations. Such loss of autonomy may discourage teachers and hinder instruction and, ultimately, student learning. Hence, grading practices present a troublesome conundrum: We want grades to be fair, but most teachers would vehemently oppose an oppressive standardization that drains enjoyment from their jobs. Thus, we continue to confound ourselves in a vain search for uniformity, misusing grades to compare students instead of simply viewing them as indicators of student progress.

In the end, if our goal is to teach students to write for an audience beyond a teacher or a rubric, we must recognize the peculiar nature of this discipline. Writing, by nature, is a personal transaction of ideas from author to readers. Our opinions of writing vary on even the most esteemed of works. Some embrace the syntactical complexity of a writer like Thomas Hardy; others view this style as a tangled impediment to the expression of ideas. Some enjoy the sarcastic humor of David Sedaris, while others would prefer a more straightforward, less sardonic view of our world. True, most of our students will not achieve the literary acclaim of a Hardy or Sedaris. In recognizing this, some teachers seek to imprison their students' writing inside the confines of sterile structures and conventions. (I once heard a colleague tell students that "when they are published, they can use a sentence fragment for effect.") Others disregard the long odds against teaching the next F. Scott Fitzgerald, and allow more leeway for students to experiment and find a voice. In my experience, those who are confined by teachers and grades and fear, learn to loathe writing and avoid doing so, defeating the purpose of memorizing rules of grammar and standards of a five-paragraph theme they will never write. The rest may never write a novel that appears on a professor's syllabus or even write an article for the local entertainment magazine. But they will write.

REFERENCES

Betts, E. (1946). Foundations of reading instruction. New York: American Book Co.

Conference on College Composition and Communication (2006). *Writing assessment: A position statement.* National Council of Teachers of English: Urbana, IL. Retrieved October 13, 2011 from http://www.ncte.org/cccc/resources/positions/writingassessment

Culham, R. (1995). *6 + 1 traits of writing.* New York: Scholastic.

Elbow, P. (1998). *Writing with power* (2nd Ed). New York: Oxford University Press.

Elbow, P. (2000). *Everyone can write: Essays toward a hopeful theory of writing and teaching writing.* New York: Oxford University Press.

Emig, J. (1971). *The composing processes of twelfth graders.* Urbana, IL: NCTE.

Hairston, M. (1982). The winds of change: Thomas Kuhn and the revolution in the teaching of writing. *College Composition and Communication, 33*(1), 76-88.

Hargis, C.H. (1999). *Teaching and testing in reading a practical guide for teachers and parents.* Springfield, IL: Charles C. Thomas.

Hillocks, G. (1986). *Research on written composition new directions for teaching.* Urbana, IL: NCTE.

Hillocks, G. (2002). *The testing trap: How state assessments control learning.* New York: Teachers College Press.

Hillocks, G. (2005). The focus on form vs. content in teaching writing. *Research in the Teaching of English*, 40(2), 238-248.

Hillocks, G. (2006). Two Decades of Research on Teaching Writing in the Secondary Schools in the US. *L1-Educational Studies in Language and Literature, 6* (2), 29-51.

Kahn, E. A. (2000). A Case Study of Assessment in a Grade 10 English Course. *The Journal of Educational Research*, *93*(5), 276.

National Council of Teachers of English. (2003, April). *Why Is Writing So Important?* Urbana, IL: Author.

Starch, D. & Elliot, E. (1912). Reliability in grading high school work in English. *School Review, 20,* 442-457.

Scherff, L., & Piazza, C. (2005). The more things change, the more they stay the same: a survey of high school students' writing experiences. *Research in the Teaching of English, 39*(3), 271-305.

Stiggins, R. & Bridgeford, N. (1985). The ecology of classroom assessment. *Journal of Educational Measurement, 22(4)*, 271-86

Stiggins, R. (1986). Inside high school grading practices. *The Northwest Regional Educational Laboratory Report.* Portland, OR: Northwest Regional Laboratory.

Stiggins, R. (1991). Assessment literacy. *Phi Delta Kappan 72,* 534-539.

Stiggins, R. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan 77*, 238-45.

Stiggins, R. (1999). Barriers to effective student assessments. *The Education Digest, 64(6),*25-9.

Stiggins, R. (2007). Assessment through the student's eyes. *Educational Leadership, 64(8),* 22-6

Tennessee Department of Education (2007). *State Report Card.* Retrieved October 24, 2008, from http://www.k-12.state.tn.us/rptcrd07

U.S. Census Bureau. (2007). *American Community Survey.* Washington, DC: U.S. Government Printing Office.

Appendices

Appendix A: Defining the 6+1 Traits of Writing

Ideas: This trait refers to the content of the paper. Is the paper focused on a defined topic? Does it have a strong thesis with pertinent support? Are the writer's ideas/thesis meaningful?

Organization: This trait refers to the paper's structure and coherence. Does the introduction provide adequate context? Are paragraphs unified? Are ideas between paragraphs connected? Does the conclusion provide closure?

Voice: This trait refers to the ability of the writer to show an appropriate personality for the writing occasion. Does the style of writing appeal to the audience? Does it match the paper's purpose?

Word Choice: This trait refers to the writer's use of diction. Are words used precisely? Are they used correctly in a way that enhances the message? Does the writer use appropriate, mature vocabulary?

Sentence Fluency: This trait refers to sentence structure and syntactic variety. Is the writing clear? Concise? Well-shaped? Does the writer present ideas in a readable manner?

Conventions: This trait refers to the writer's use of standard English grammar. This includes avoiding errors such as subject-verb disagreement, pronoun-antecedent disagreement, run-on sentences, fragments, spelling, etc. While publishable perfection is not necessarily expected, errors should be minimal and not distract from the essay's clarity.

Presentation: This refers to the appearance and format of the paper. Does the writer conform to expectations for margins, spacing, font, title page, etc.? Is the paper's appearance appealing to the reader?

Appendix B: The Student Paper

Why Abortion Should Be Illegal
Argumentative Essay

You wait anxiously in the stark office, rocking slightly as a way to compose yourself. You remember how your dad used to cradle and rock you so that you would cease crying, if only he could be here now to comfort you. But nobody knows and that is how you prefer to keep it, at least that is how you feel for the moment. Oh, the shame you would feel if they found out! Not only would you be humiliated, but to see that look of horror on your parents' faces…merely thinking about it is insufferable. Luckily, you were able to keep it a secret since you moved out of your parents' house a year ago and have not seen them since. Because if you had, you would have broken down and told them, but now no one is the wiser.

You gaze sullenly around at the other women, most of them part of a team, a couple…a father is there to greet the supposed to be joyful news with. Tears begin to well up in your already swollen eyes; pain and hate fill your gut. You loathe men, especially *that* man and the worst part is that you don't even know who you are hating: you never saw his face, just the gun that was pointed at yours. Now, inside of you, a portion of that despicable animal is combining with part of you to create one, a baby. It makes you sick.

The nurse calls your name and summons you through the sanitized halls of the building into a slightly more cheerful examination room. Listlessly you follow. She leaves you alone with your poisoning thoughts: this baby will end the life that you once knew…now everything will revolve around this child. You are only 19 and your whole life remains ahead of you, college, a career, a *family*. You do not even have a boyfriend; how are you supposed to get one now that you are pregnant? And most importantly, what will your family think of you? Would they help you support it? As a college student you are financially unstable and completely unable to support a baby…Then, almost an hour later, a doctor with a gentle yet shockingly placid voice brings you back to reality, back to life, with the option of abortion. It is not that you have never considered abortion, but now it is an *option*; someone else is presenting this alluring idea to you. But should you? If you were to carry and deliver this child there would be no way to conceal it, even if you were to give it up for adoption. But could you be so selfish as to kill a baby?

Abortion is such a loaded word that maybe it is overlooked, but under the surface it is not about freedom from a life with the responsibilities of a baby but about consequences. An abortion can be a seemingly effortless way to escape from your "problems" but in actuality it causes the death of a child, and much grief. Not to mention other complications from the procedure, such as a torn cervix. Regardless, abortion is alarmingly common and many women, such as the one in the aforementioned story, confront the decision to abort their baby daily. The Alan Guttmacher Institute, an organization that is a global leader in sexual and reproductive studies, states that, "Worldwide, the lifetime average of abortion is about 1 per woman." These women are now in danger of the adverse effects of abortion, such as post-traumatic stress disorder. Author and Feminist, Frederica Mathewes-Green, claims, "Pro-life and pro-choice can agree: abortion is a tragedy, and women deserve better choices." Essentially, abortion

should be illegal because it causes negative physical and emotional distress on the recipient and her family.

The most predominant issues that must be addressed are the physical complications that accompany abortions. David C. Reardon, Ph D. and director of the Elliot Institute, compiled a list of data for the Ellis Institute, a non-profit corporation that performs research on the impact of abortion. He states, "Approximately 10% of women

undergoing elective abortion will suffer immediate complications, of which approximately one-fifth (2%) are considered life threatening." The Silent No More Awareness Campaign claims, "In the US, over 140,000 women a year have immediate complications from abortion." In a packet given to women considering abortion by the Pregnancy Help Center of Knoxville claims, "Even though abortion is legal, it is not safe. The standard of care to protect women's health is often inadequate and some abortionists move from state to state as a way to avoid investigation and patient complaints." The packet proceeds saying, "Most abortionists do not screen for risk factors or determine whether abortion will benefit their patients. Proper screening would eliminate 70% or more of all abortions."

The most common complications that occur at the time of an abortion are: infection, excessive bleeding, blockage of an artery, a painfully inflamed abdomen caused by a perforation of the uterus, anesthesia complications, convulsions, hemorrhage, cervical injury (which causes an increased chance of miscarriage), endotoxic shock (a condition that leads to low blood pressure and decreased blood flow), second degree burns, chronic abdominal pain, vomiting, gastro-intestinal disturbances, and Rh sensitization. Rh sensitization can occur when a woman with Rh-negative blood is exposed to blood from her Rh-positive fetus. Once the mother is exposed to Rh-positive blood, her immune system produces antibodies that can destroy the fetus's Rh-positive red blood cells. But not only does abortion have immediate consequences it also increases your chances of contracting other complications. In fact, "The risk of breast and cervical cancer almost doubles after one abortion, and rises even further with two or more abortions" (Reardon). The truth is that abortions are harmful and possibly life threatening.

Another aspect that must be considered is the spiritual one. Abortion takes an immense emotional toll on women and their family. The Elliot Institute asserts,

In a study of post-abortion patients only 8 weeks after their abortion, researchers found that 44% complained of nervous disorders, 36% had experienced sleep disturbances, 31% had regrets about their decision, and 11% had been prescribed psychotropic medicine by their family doctor". One of the most common side effects is Post-Traumatic Stress Disorder (PTSD), also referred to as Post-Abortion Trauma. PTSD is a psychosomatic dysfunction caused by a traumatic experience which floods a person's normal defense mechanisms. PTSD results in intense fear, feelings of helplessness, being trapped, and loss of control." Knoxville's Pregnancy Help Center lists the symptoms as: "bouts of crying, depression, guilt, intense grief, rage, emotional numbness, anxiety, flashbacks, sleep disturbances, suicidal urges, and discomfort around babies or pregnant women.

Abortion is also linked with a fifty percent increase in risk of alcohol and or drug abuse among women because they cannot find any other way to cope with their feelings. "Researchers in Finland have identified a strong association between abortion and suicide in a records based study; approximately 60 percent of women who experience post-abortion report suicidal ideation, with 28 percent actually attempting suicide, of which half attempted suicide two or more times" (Reardon). If a woman is suffering from guilt related to the abortion there is likely to be reduced maternal bonding with future children. Subsequently, those women are more likely to neglect and or abuse their other children. Yet, a woman's guilt, or other symptoms for that matter, would not just affect her potential children; her spouse or other closely related persons would also be adversely affected by her pain.

In conclusion, abortion should be illegal because of the effects that abortion can have on the woman receiving one. The Pregnancy Health Center of Knoxville Tennessee states, "On average, there is an 80% increase in doctor visits and a 180% increase in doctor visits for psychosocial reasons after abortion." Pro-Choice advocators intend to give women a choice to their lives, a chance to live without a baby, but by giving those women that one choice they are stripping a child of a lifetime of choices and they are also ignoring the consequences that abortion has on the woman and anyone connected to her.

Works Cited

Reardon, David C. "A List of Major Psychological Sequelae of Abortion." 1997. Elliot Institute. <www.Afterabortion.org>.

Reardon, David C. "A List of Major Physical Sequelae Related to Abortion." 1997. Elliot Institute.

<http://www.abortionfacts.com/reardon/effect_of_abortion.asp>.

Forney, Georgette. "Ten Facts that Women Need to Know about Abortion." 2002.

Silent No More Awareness Campaign. http://www.silentnomoreawareness.org/about/.

Reardon, David C. "Research on Post-Abortion Issues." 1997. Elliot Institute. www.afterabortion.org.

Pregnancy Help Center. "Key Facts about Abortion."

Pregnancy Help Center. "Risks of Abortion."

Appendix C: Distribution of Scores on a 100-Point Scale

Score	Freq	Percent	Cum Percent
50	1	1.4	1.4
56	1	1.4	2.7
63	1	1.4	4.1
65	2	2.7	6.8
66.67	1	1.4	8.2
70	5	6.8	15.1
72	2	2.7	17.8
74	2	2.7	20.5
75	5	6.8	27.4
76	3	4.1	31.5
77	1	1.4	32.9
78	2	2.7	35.6
79	2	2.7	38.4
80	6	8.2	46.6
81	1	1.4	47.9
82	1	1.4	49.3
83	3	4.1	53.4
84	6	8.2	61.6
85	4	5.5	67.1
86	1	1.4	68.5
87	3	4.1	72.6
88	1	1.4	74.0
89	1	1.4	75.3
90	6	8.2	83.6
91	1	1.4	84.9
92	1	1.4	86.3
93	5	6.8	93.2
94	1	1.4	94.5
95	3	4.1	98.6
96	1	1.4	100.0
Total	73	100.0	

Citation:

Brimi, Hunter M. (2011). Reliability of Grading High School Work in English. *Practical Assessment, Research & Evaluation*, 16(17). Available online: http://pareonline.net/getvn.asp?v=16&n=17

Author:

Hunter M. Brimi, English Teacher
Farragut High School
Knoxville, TN
Hbrimi [at] aol.com